REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
07-11-2013	Final Report	23-Sep-2008 - 22-Sep-2012

4. TITLE AND SUBTITLE
Collaborative and Persistent Target Tracking and Acquisition (54348-CI)

5a. CONTRACT NUMBER
W911NF-08-1-0504

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHORS
Ying Wu

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Northwestern University Chicago Campus
Office of Sponsored Research
633 Clark St., Rebecca Crown Bldg. 2-502
Evanston, IL 60208 -1110

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS (ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
54348-CS.22

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Extending sensing capacity in target acquisition and tracking is very important for Army's future force. Video sensors provide both opportunities and challenges. One on hand, they bring rich and detailed information to facilitate identifying and recognizing the targets. But on the other hand, such visual information is subject to large uncertainties and ambiguities induced by factors such as cluttered and distractive backgrounds, illumination changes, visual occlusion, and low image quality. All these difficulties have impeded the development of effective

15. SUBJECT TERMS
Target Tracking, Video Analytics

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Ying Wu
UU	UU	UU	UU		19b. TELEPHONE NUMBER 847-491-2901

Report Title

Collaborative and Persistent Target Tracking and Acquisition
(54348-CI)

ABSTRACT

Extending sensing capacity in target acquisition and tracking is very important for Army's future force. Video sensors provide both opportunities and challenges. One on hand, they bring rich and detailed information to facilitate identifying and recognizing the targets. But on the other hand, such visual information is subject to large uncertainties and ambiguities induced by factors such as cluttered and distractive backgrounds, illumination changes, visual occlusion, and low image quality. All these difficulties have impeded the development of effective and robust methods for persistent target tracking and acquisition in unconstrained environments. As a result, it is still challenging while desirable to have long-duration target tracking systems in many emerging applications, e.g., video surveillance and sensor networks.

The goal of the proposed project is to overcome the challenges that confront persistent target tracking and acquisition, by developing new and effective computational models and methods that bridge computational sensing and some aspects in human perception of visual dynamics, and by creating multi-level synergy among various modalities and sensors.

Enter List of papers submitted or published that acknowledge ARO support from the start of the project to the date of this printing. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Received		Paper
11/07/2013	1.00	N. Jiang, H. Su, W. Liu, Y. Wu. Discriminative Metric Preservation for Tracking Low Resolution Targets, IEEE Transactions on Image Processing, (03 2012): 1284. doi: 10.1109/TIP.2011.2167345
11/07/2013	3.00	Nan Jiang, Wenyu Liu, Ying Wu. Learning Adaptive Metric for Robust Visual Tracking, IEEE Transactions on Image Processing, (08 2011): 2288. doi: 10.1109/TIP.2011.2114895
11/07/2013	19.00	Heng Su, Jie Zhou, Ying Wu. Super-resolution without Dense Flow, IEEE TRANSACTIONS ON Image Processing, (04 2012): 1782. doi:
11/07/2013	20.00	Ming Yang, Zhimin Fan, Jialue Fan, Ying Wu. Tracking Non-stationary Visual Appearances by Data-driven Adaptation, IEEE TRANSACTIONS ON Image Processing, (07 2009): 1633. doi:
11/07/2013	21.00	Ming Yang, Ying Wu, Gang Hua. Context-Aware Visual Tracking, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,, (07 2009): 1195. doi:
11/07/2013	16.00	Xiaohui Shen, Ying Wu, Jialue Fan. What are We Tracking: A Unified Approach of Tracking and Recognition, IEEE TRANSACTIONS ON Image Processing, (02 2013): 549. doi:
11/07/2013	17.00	Jialue Fan, Xiaohui Shen, Ying Wu. Scribble Tracker: A Matting-based Approach for Robust Tracking, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE,, (08 2012): 1633. doi:

TOTAL: **7**

Number of Papers published in peer-reviewed journals:

(b) Papers published in non-peer-reviewed journals (N/A for none)

Received Paper

TOTAL:

Number of Papers published in non peer-reviewed journals:

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received Paper

TOTAL:

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Peer-Reviewed Conference Proceeding publications (other than abstracts):

Received		Paper
11/07/2013	2.00	Wenyu Liu, Ying Wu, Nan Jiang. Adaptive and discriminative metric differential tracking, 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 19-JUN-11, Colorado Springs, CO, USA. : ,
11/07/2013	4.00	Nan Jiang, Wenyu Liu, Heng Su, Ying Wu. Tracking low resolution objects by metric preservation, 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 19-JUN-11, Colorado Springs, CO, USA. : ,
11/07/2013	5.00	Ming Yang, Ying Wu, Junsong Yuan. Mining discriminative co-occurrence patterns for visual recognition, 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 19-JUN-11, Colorado Springs, CO, USA. : ,
11/07/2013	6.00	Philip Lee, Ying Wu. Nonlocal matting, 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 19-JUN-11, Colorado Springs, CO, USA. : ,
11/07/2013	7.00	Jiang Wang, Zhuoyuan Chen, Ying Wu. Action recognition with multiscale spatio-temporal contexts, 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 19-JUN-11, Colorado Springs, CO, USA. : ,
11/07/2013	8.00	Xiaohui Shen, Ying Wu. A Unified Approach to Salient Object Detection via Low Rank Matrix Recovery, IEEE Conf. on Computer Vision and Pattern Recognition. 20-JUN-12, . : ,
11/07/2013	10.00	Nan Jiang, Wenyu Liu, Ying Wu. Order Determination and Sparsity-Regularized Metric Learning for Adaptive Visual Tracking, IEEE Conf. on Computer Vision and Pattern Recognition. 16-JUN-12, . : ,
11/07/2013	9.00	Jiang Wang, Ying Wu, Zhuoyuan Chen. Decomposing and Regularizing Sparse/Non-sparse Components for Motion Field Estimation, IEEE Conf. on Computer Vision and Pattern Recognition. 16-JUN-12, . : ,
11/07/2013	11.00	Xiaohui Shen, Ying Wu. Sparsity Model for Robust Optical Flow Estimation at Motion Discontinuities, IEEE Conf. on Computer Vision and Pattern Recognition. 20-JUN-10, . : ,
11/07/2013	12.00	Jialue Fan, Ying Wu, Shenyang Dai. Discriminative Spatial Attention for Robust Tracking, European Conf. on Computer Vision. 20-SEP-10, . : ,
11/07/2013	13.00	Jialue Fan, Xiaohui Shen, Ying Wu. Closed-loop Adaptation for Robust Tracking, European Conf. on Computer Vision. , . : ,
11/07/2013	14.00	Xiaohui Shen, Ying Wu. Exploiting Sparsity in Dense Optical Flow, IEEE Int'l Conf. on Image Processing. 15-OCT-10, . : ,
11/07/2013	15.00	Jiang Xu, Ying Wu, Jialue Fan. Context-aware Tracking of Small Targets in Video, SPIE Symposium on Optical Engineering and Applications. 25-AUG-09, . : ,

TOTAL: **13**

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts):

(d) Manuscripts

Received   Paper

TOTAL:

Number of Manuscripts:

Books

Received   Paper

TOTAL:

Patents Submitted

Patents Awarded

Awards

Graduate Students

NAME	PERCENT SUPPORTED	Discipline
Jialue Fan	0.50	
Xiaohui Shen	0.50	
Nan Jiang	0.10	
Jiang Xu	0.30	
FTE Equivalent:	**1.40**	
Total Number:	**4**	

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Ying Wu	0.10	
FTE Equivalent:	**0.10**	
Total Number:	**1**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period: 0.00

The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:...... 0.00

Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):...... 0.00

Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:...... 0.00

The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense 0.00

The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:...... 0.00

Names of Personnel receiving masters degrees

NAME	
Total Number:	

Names of personnel receiving PHDs

NAME	
Jiang Xu	
Jialue Fan	
Xiaohui Shen	
Nan Jiang	
Total Number:	**4**

Names of other research staff

NAME	PERCENT_SUPPORTED
Heng Su	0.05
FTE Equivalent:	**0.05**
Total Number:	**1**

Sub Contractors (DD882)

Inventions (DD882)

Scientific Progress

In the first period of the project (9/2008-2/2009), we mainly investigated the selective attention model that incorporates four computational processes: early attentional selection, collaborative tracking, context-aware learning for late selection, and robust information fusion. Details can be found in the previous year report.

In the second period of the project (2/2009-2/2010), we focused our study on context-aware tracking and robust multimodal fusion. Based on our new model of context flow, we developed a powerful method to track general targets based on dense context, and to track small targets based on sparse context.

In the third period of the project (2/2010-2/2011), we investigated how to embed discriminative learning in the late selection process, and how the higher-level selection process can influence the lower-level selection process.

In the fourth period of the project (2/2011-2/2012), we studied three important issues in the proposal SSA model. (1) The most critical issue in visual tracking is to match the target over time. Reliable matching depends on the choice of appropriate similarity metric. We study how the metric can be automatically learnt and adapted for selective attention. (2) A practical problem in real applications is to track small targets in low-resolution video. We propose an innovative approach to track low-resolution targets without explicitly performing video super resolution. (3) The combination of small patterns can be more discriminative. We study a new approach to automatically discover disjunctive and conjunctive patterns from data. This pattern mining approach is one of the core components in the proposed SSA model.

In the last period of the project (2/2012-9/2012), we kept investigating the important issues in online metric adjustment and selective attention, as identified in the past years. Specifically, we have studied (1) the optimal order determination in metric learning, and (2) the computational modeling on the unification of spatial selection and attribute selection.

(1) Order Determination for the Optimal Metric

Our studies in the previous years in this project demonstrated that (a) the matching metric is one of the most critical issues in target tracking, and using a pre-defined metric is not plausible; (b) the matching metric needs to be adaptive to the tracking scene so as to better separate the target and the distracters, and (3) such scene-aware metric can be learned and adapted along with target tracking, if good training data can be collected on-the-fly with the scene change, e.g., based on some prior knowledge. In the past several years, we have developed two innovative approaches to learn adaptive metric for tracking.

We also found that the order of the metric plays an important role. Adjusting the matching metric can be regarded as projecting or mapping the original feature space to a new feature space (linearly or nonlinearly). The order of the metric is the dimension of such a new feature space. If a lower dimension is used, the learned metric may be an under-fit of the data; and a higher dimension generally leads to over-fit of the data. Therefore, we study this important issue of order determination, i.e., finding the optimal dimension for the projected feature space. In the linear case, it is the rank of the Mahalonobis metric. This is important to the accuracy and the generalizability of the learned metric. But this is a very challenging problem, as the order is a structural parameter, and it is not differentiable.

We have found a novel solution to this structural order determination problem. The idea is to introduce sparsity prior to regularize metric learning. The purpose is to find the lowest rank of the projection while achieving the best classification. The introduced sparsity regularization serves this purpose, and leads to the lowest possible dimension. This can actually be viewed as a minimum description length regularization for metric learning. We have validated this new solution on standard benchmark datasets, and demonstrated its effectiveness in visual target tracking tasks.

(2) Unification of Spatial Selection and Attribute Selection

One of the most challenging issues that prevent video-based tracking from being reliable and useful in real applications is visual distracters. Visual distracters are those that have similar visual appearance to the target, but are actually false positives. They mislead and fail the tracker. This is very common when tracking low-resolution targets, and it is almost universal in real application scenarios.

The resilience of a good tracker depends not only one the matching metric, but also on the characteristics of the target region to be matched. The two tasks, i.e., learning the best metric and selecting the distracter-resilient target regions, actually correspond to the attribute selection and spatial selection processes in the human visual perception. Spatial selection chooses some spatial salient regions for matching, and attribute selection constructs and forms some good features for matching. Although there have been many studies on these two topics separately in the literature, it is desirable to have a unified approach to integrate both in a coherent framework.

We have been making an initial attempt to unify the modeling of these two tasks for an effective solution, based on the introduction of a new quantity called soft visual margin. As a function of both matching metric and spatial location, it measures the discrimination between the target and its spatial distracters, and characterizes the reliability of matching. We have obtained a novel method to jointly determine the best spatial location and the optimal metric. Based on that, a solid distracter-resilient region tracker is designed, and its effectiveness is validated and demonstrated through extensive experiments.

Technology Transfer

Collaborative and Persistent Target Tracking and Acquisition
54348-CI

Ying Wu

Electrical Engineering and Computer Science
Northwestern University, Evanston, IL 60208

1. Objective

Extending sensing capacity in target acquisition and tracking is very important for Army's future force. Video sensors provide both opportunities and challenges. One on hand, they bring rich and detailed information to facilitate identifying and recognizing the targets. But on the other hand, such visual information is subject to large uncertainties and ambiguities induced by factors such as cluttered and distractive backgrounds, illumination changes, visual occlusion, and low image quality. All these difficulties have impeded the development of effective and robust methods for persistent target tracking and acquisition in unconstrained environments. As a result, it is still challenging while desirable to have long-duration target tracking systems in many emerging applications, e.g., video surveillances and sensor networks.

The goal of the proposed project is to overcome the challenges that confront persistent target tracking and acquisition, by developing new and effective computational models and methods that bridge computational sensing and some aspects in human perception of visual dynamics, and by creating multi-level synergy among various modalities and sensors.

2. Approach

We propose novel research of collaborative and persistent target tracking and acquisition, by creating a multi-level synergism to address the difficulties mentioned above including visual clutters and occlusions in video-based tracking, robustness and persistency, target re-acquisition and association in distributed video sensors, and information inconsistency in fusion.

- The *synergetic selective attention model* (SSA). Motivated by some psychological findings, we propose a new computational model for collaborative and persistent target acquisition and tracking. This new model has four components: early attentional selection, collaborative tracking, context-aware learning and late selection, and robust information integration. This new model is fundamentally different from the existing models for target tracking. A multi-level synergism is enabled in a coherent way based on this new computational model.

- *Synergy in selective visual attention* to handle visual clutters and occlusions. Based on the SSA model, the collaboration among the selected attentional regions leads to the resilience to visual clutters and occlusions.

- *Synergy in contextual visual analysis* for robustness and persistency in visual tracking. This SSA model accommodates the visual contexts of the target of interest. Context-aware learning in the proposed approach produces robust and persistent visual tracking.

- *Synergy in multiple distributed cameras* for target re-acquisition and association. The visual selection and attention mechanism in the SSA model also allows target re-acquisition and association in a camera network.

- *Synergy in robust information fusion* for integrating multiple sources and multiple modality sensors. The robust information integration mechanism in the SSA model is able to handle the information inconsistency among multiple modality sensors.

One important innovation of this project is the introduction of some important findings in physiological and psychological vision studies to the computational models for robust and persistent target tracking and acquisition. To the best of our knowledge, this research is a pioneering attempt of its kind. Most existing video-based methods has been confronted by many challenges in practice, as they have over simplified the complicated process in the human perception of visual dynamics, and have ignored many powerful mechanisms in the human visual perception, such as *selective attention* and *context-awareness*. The proposed new computational models is the first of its kind in incorporating selective attention and context-awareness in target tracking, and is expected to result in more powerful tracking methods.

Another innovation is the *multi-level synergism* that integrates the analysis of sensory data at four levels. Firstly, we represent the visual appearance of the target by its local invariant regions (called attentional regions). The synergy at the visual region level enables selective visual attention and overcomes the difficulties of cluttered backgrounds and partial occlusions. Secondly, the tracking of the target is not only concentrated on the target itself, but also incorporating other objects in its spatial context. Such a contextual synergy at the object level greatly enhances the robustness and persistence for long-duration tracking. Thirdly, the synergy in distributed cameras facilitates robust data association and target re-acquisition. Last, the synergy in multi-modal sensors enables robust information fusion for more accurate target tracking and acquisition. This unification has never been seen in the literature.

3. Scientific Barriers

The major challenge in video-based target tracking and acquisition lies in the large uncertainties and ambiguities in video data. The specific scientific barriers in this research are in the following four aspects:

- *Uncertainties induced by visual clutters and occlusions.* The major hurdles in video-based target tracking include the background clutters and occlusions. Clutters generate false positive matches that distract the tracker, and occlusions conceal the visual observations for matching. It is still a very challenge task to detect and locate the target in a clutter and to handle occlusions.

- *Robustness and persistence in long-duration target tracking.* This is not trivial, as the target's visual appearance may undergo enormous and unexpected changes due to the changes of view, pose, scale, illumination and the interferences from the environment. It is desirable to have truly intelligent and adaptive approaches.

- *Target re-acquisition and association in a network of video sensors*. It is important to know when and where the targets have been seen in different cameras, i.e., to associate their tracks and to re-acquire them in other cameras. This problem has not been well investigated yet.

- *Inconsistency in multi-modal sensor fusion*. Beside video sensors, other imaging and non-imaging sensors may be used. Integrating multi-modal sensors may enhance the tracking accuracy, but in practice it is often confronted by a special difficulty that different sensors may provide inconsistent or even conflicting information. Regardless fusion leads to meaningless and wrong results. Detecting and handling the inconsistency in fusion is still an open research problem.

4. Significance

The proposed research in this project will have a significant potential impact on Army in the following ways:

- In many military scenarios, surveillance is important for both battlefields and non-battlefields. Good surveillance systems provide soldiers safe and secure environments. The proposed target tracking and acquisition methods will greatly improve the performance of video-based surveillance systems;

- Digital Soldier is a trend for the future warfare, where an individual soldier's abilities in sensing and communication are largely extended. It can be envisioned that such a Digital Soldier is equipped with a powerful and intelligent video sensor, which is able to see what the soldier is not able to see or does not have time to see. This *third eye* will be very significant for the survivability of the solider in action. This proposed target tracking and acquisition method can be a critical and fundamental component for this *third eye*;

- Military robots, such as combat robots, are becoming part of the battlefield. Some of them are autonomous, and they need powerful sensing capacity to move and act. For example, accurate and real-time targeting is important for saving the ammunitions of a combat robot. They are beneficial from the proposed research as well;

- Urban warfare is challenging, as the environment is complex. The proposed methods of persistent target acquisition and tracking can be quite useful in urban warfare. For example, distributed cameras will collaborate to locate and track the target in a local area. Such real-time information can be utilized for further actions.

5. Accomplishments of Current Period

In the first period of the project (9/2008-2/2009), we mainly investigated the selective attention model that incorporates four computational processes: early attentional selection, collaborative tracking, context-aware learning for late selection, and robust information fusion. Details can be found in the previous year report.

In the second period of the project (2/2009-2/2010), we focused our study on context-aware tracking and robust multimodal fusion. Based on our new model of context flow, we developed a powerful method to track general targets based on dense context, and to track small targets based on sparse context.

In the third period of the project (2/2010-2/2011), we investigated how to embed discriminative learning in the late selection process, and how the higher-level selection process can influence the lower-level selection process.

In the fourth period of the project (2/2011-2/2012), we studied three important issues in the proposal SSA model. (1) The most critical issue in visual tracking is to match the target over time. Reliable matching depends on the choice of appropriate similarity metric. We study how the metric can be automatically learnt and adapted for selective attention. (2) A practical problem in real applications is to track small targets in low-resolution video. We propose an innovative approach to track low-resolution targets without explicitly performing video super resolution. (3) The combination of small patterns can be more discriminative. We study a new approach to automatically discover disjunctive and conjunctive patterns from data. This pattern mining approach is one of the core components in the proposed SSA model.

In the last period of the project (2/2012-9/2012), we kept investigating the important issues in online metric adjustment and selective attention, as identified in the past years. Specifically, we have studied (1) the optimal order determination in metric learning, and (2) the computational modeling on the unification of spatial selection and attribute selection.

5.1 Order Determination for the Optimal Metric

Our studies in the previous years in this project demonstrated that (a) the matching metric is one of the most critical issues in target tracking, and using a pre-defined metric is not plausible; (b) the matching metric needs to be adaptive to the tracking scene so as to better separate the target and the distracters, and (3) such scene-aware metric can be learned and adapted along with target tracking, if good training data can be collected on-the-fly with the scene change, e.g., based on some prior knowledge. In the past several years, we have developed two innovative approaches to learn adaptive metric for tracking.

We also found that the order of the metric plays an important role. Adjusting the matching metric can be regarded as projecting or mapping the original feature space to a new feature space (linearly or nonlinearly). The order of the metric is the dimension of such a new feature space. If a lower dimension is used, the learned metric may be an under-fit of the data; and a higher dimension generally leads to over-fit of the data. Therefore, we study this important issue of order determination, i.e., finding the optimal dimension for the projected feature space. In the linear case, it is the rank of the Mahalonobis metric. This is important to the accuracy and the generalizability of the learned metric. But this is a very challenging problem, as the order is a structural parameter, and it is not differentiable.

We have found a novel solution to this structural order determination problem. The idea is to introduce sparsity prior to regularize metric learning. The purpose is to find the lowest rank of the projection while achieving the best classification. The introduced sparsity regularization serves this purpose, and leads to the lowest possible dimension. This can actually be viewed as a minimum description length regularization for metric learning. We have validated this new solution on standard benchmark datasets, and demonstrated its effectiveness in visual target tracking tasks.

5.2 Unification of Spatial Selection and Attribute Selection

One of the most challenging issues that prevent video-based tracking from being reliable and useful in real applications is visual distracters. Visual distracters are those that have similar visual appearance to the target, but are actually false positives. They mislead and fail the tracker. This is very common when tracking low-resolution targets, and it is almost universal in real application scenarios.

The resilience of a good tracker depends not only one the matching metric, but also on the characteristics of the target region to be matched. The two tasks, i.e., learning the best metric and selecting the distracter-resilient target regions, actually correspond to the attribute selection and spatial selection processes in the human visual perception. Spatial selection chooses some spatial salient regions for matching, and attribute selection constructs and forms some good features for matching. Although there have been many studies on these two topics separately in the literature, it is desirable to have a unified approach to integrate both in a coherent framework.

We have been making an initial attempt to unify the modeling of these two tasks for an effective solution, based on the introduction of a new quantity called soft visual margin. As a function of both matching metric and spatial location, it measures the discrimination between the target and its spatial distracters, and characterizes the reliability of matching. We have obtained a novel method to jointly determine the best spatial location and the optimal metric. Based on that, a solid distracter-resilient region tracker is designed, and its effectiveness is validated and demonstrated through extensive experiments.

6. Collaborations and Leveraged Funding

This project leverages the research of an NSF project, "*CAREER: Visual Analysis of High-Dimensional Motion: A Distributed/Collaboration Approach*" 1/2004-2/2011. This NSF project is focused on the estimation of complex and high-dimensional motions, such as the articulated motion of the human body, and the deformable motion of elastic shapes.

7. Conclusions

Persistent target acquisition and tracking in video is one of the key issues in video analysis and understanding, and thus it is a fundamental component in many emerging applications. Although video convey rich information about the scene, the uncertainties in the visual information induced by cluttered background and occlusion largely make this problem quite challenging. This project is targeted on developing new and effective computational models and methods that bridge computational sensing and some aspects in human perception of visual dynamics, especially selective attention and context-awareness in human visual perception. This project aims to create a multi-level synergism to address the difficulties including visual clutters and occlusions in video-based tracking, robustness and persistency, target re-acquisition and association in distributed video sensors, and information inconsistency in fusion. This project shall lead to a new computational model, a class of effective methods, and a set of powerful tools for video-based target tracking and acquisition.

8. Technology Transfer

We have not had technology transfer activities so far, but we plan to develop solid collaborations with researchers at ARL as well as some industrial companies.

9. Future Research Plans

Although the project has concluded with a promising new computational model for persistent target tracking and a family of robust tracking methods, the research can be advanced further on the following directions:

- This computational model is motivated by some interesting findings in the studies of the human visual perception. It attempts to give a plausible explanation to the known but vague mechanisms in human visual perception. A very interesting and far-reaching research is to investigate how this computational model can be used in the visual cognition and psychological studies, to predict some hidden mechanisms in human visual cognition. For example, what aspects of selection should be in the early stage, and what aspects are learned in the latter stage. This shall largely bridge the study of computational vision and human vision.

- Another direction is hardware implementation of this computational models and methods. Some components in the proposed SSA model demand more computational resources, e.g., the online visual pattern mining component and the online metric learning component. As they need to be adaptive and updated over time, it is difficult for our current software-based implementation to achieve real-time performance. It is of great interest if these methods can be parallelized and have hardware implementations.

10. Anticipated Scientific Accomplishments

The following is what we anticipate on this research project:

- **A new computational model** for video-based target tracking and acquisition. This new model incorporates the selective attention mechanism of the human visual perception into the mathematical modeling, and it will enable context-aware motion tracking, and unify the adaptation at various levels of visual information.

- **Effective methods** for target tracking and acquisition. Our new computational model leads to a class of effective target tracking methods that are able to automatically focus their computational attention to selected image regions and features, and able to make user of the contextual information from the environments. We expect this class of new tracking method to significantly outperform the methods in the current state-of-the-art, in obtaining robust and persistent target tracking for a long duration.

- **Powerful tools** for video analysis and target tracking. In this project, a set of powerful tools will be developed, including but not limited to contextual flow computation, salient attentional region determination, on-line mining of motion correlation, learning dynamic random field, detection of information inconsistency, etc. These powerful tools are useful for a large variety of video analysis tasks.